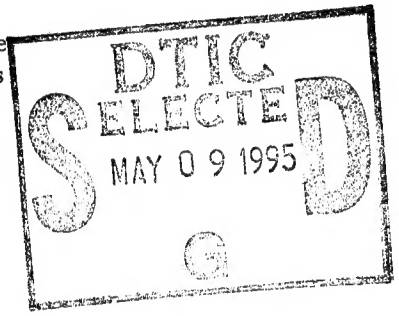
ACCELERATED LIFE TESTS THAT MAXIMIZE SHANNON INFORMATION

by

Nick Polson
Nozer D. Singpurwalla
Isabella Verdinelli

GWU/IRRA/Serial TR-90/2

The George Washington University
School of Engineering and Applied Science
Institute for Reliability and Risk Analysis
Washington, DC 20052



Research Supported by

Contract N00014-85-K-0202
Project NR 042-372
Office of Naval Research

and

Grant DAAL03-87-K-0056
U.S. Army Research Office

Accelerated life tests that maximise Shannon information

By Nick Polson, Nozer Singpurwalla and Isabella Verdinelli[1]

Abstract

The current literature on accelerated life tests emphasizes issues of inference and extrapolation under a given design, rather than the optimal design of an accelerated test. In this expository paper, we outline a framework for a coherent approach for the conduct of accelerated tests. Our approach is based on the Bayesian paradigm for experimental designs under linear models and involves the specification of a utility function. For the latter we adopt Shannon Information between predictive densities. The optimal design is then selected via the principle of maximum expected utility. We illustrate the approach with some special cases.

Key words and phrases. Bayesian Methods, Biometry, Kullback-Leibler Distance, Linear Models, Clinical Trials, Dose Response Experiments, Overstress Tests, Power Law, Reliability.

[1]Nick Polson is Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pa. 15213. Nozer D. Singpurwalla is Professor, Department of Operations Research and Statistics, The George Washington University and Isabella Verdinelli is Associate Professor, Departimento di Statistica, Università di Roma. This work was carried out when Singpurwalla and Verdinelli were visiting CMU.

1 Introduction

Let $S_1, S_2, \ldots,$ be a collection of environmental conditions (stresses in engineering, doses in biometry) at which it is possible to conduct life tests, and suppose that $S_1 \prec S_2 \prec \ldots$, where "$S_i \prec S_j$" denotes the fact that S_j is judged to be more severe than S_i. Let $S_u \prec S_1$, denote an environmental condition at which it may or may not be possible to conduct a life test; S_u will be referred to as the **use conditions** environment. Let X_i denote the life length of an item under S_i, and suppose that whenever $S_i \prec S_j$, $X_i \geq^{st} X_j$, where "$X \geq^{st} Y$" denotes the fact that $P(X \geq x) \geq P(Y \geq x)$ for all $x \geq 0$. Suppose X_i is assessed to have distribution $F_i(\cdot \mid \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is a vector of unknown parameters. It is typical [Mann, Schafer and Singpurwalla (1974)] to let $F_i(\cdot \mid \boldsymbol{\theta}_i) = F(\cdot \mid \boldsymbol{\theta}_i)$, for all i, and to assume that F is a commonly used distribution function in life testing, such as an exponential, a gamma, a lognormal or a Weibull.

The term "accelerated test" refers to the fact that the conditions under which testing is done are more severe than those under which the item is intended to be used, so that the failure time of an item is shortened. The dual scenario, namely testing under conditions less severe than use conditions, has received less attention because the two are conceptually equivalent. The aim of accelerated testing is to facilitate an early inference about X_u bearing in mind the requirement that $X_u \geq^{st} X_1$. To accomplish the above, one embarks on an experiment, called an **accelerated life test**, and makes some assumptions which facilitate an analysis of the data from the experiment. Of the assumptions that are made, a key one is the **time transformation function**, whose aim is to relate $\boldsymbol{\theta}_i$ and S_i, $i = 1, 2, \ldots,$. The design of an accelerated life test calls for a specification of several quantities, namely the k levels of testing, say $S_1, < S_2 < \ldots < S_k$, $k \geq 1, n_i$, the number of items to be tested under S_i, r_i, the number of failures to be observed under S_i, and/or T_i, the time of termination of the life test under S_i. Clearly, if X_{ij} denotes the time to failure of the j-th item

under S_i, then $X_{ij} \leq T_i$, $I = 1, \ldots, k$, and $j = 1, \ldots, r_i \leq n_i$. Other protocols for the design would involve a combination of the above, such as testing until $\min(T_i, X_{ir_i})$, or testing under S_i for a duration T_i, and then continuing to test the unfailed items under S_{i+1} for a duration T_{i+1}, and so on. The latter tests are called **step-stress tests** [Shaked and Singpurwalla (1983), DeGroot and Goel (1979)], and will not be the topic of discussion here.

To date, much of the statistical literature on accelerated life testing has focussed on issues of inference about X_u given an accelerated life testing protocol, rather than the more encompassing problem of designing an optimal protocol; see, for example, Meinhold and Singpurwalla (1987), Blackwell and Singpurwalla (1988) for recent bibliographic citations. The few published papers which address the question of design appear to lack a sound foundation and consequently fail to recognize the important role played by the time transformation function in developing an optimal design. The aim of this paper is to prescribe a framework for the coherent conduct of an accelerated life test. Specifically, what is needed is a paradigm for specifying a choice of k and the associated stress levels $S_1, \ldots, S_k$, the number of items to be tested $n_1, \ldots, n_k$, the test truncation times $\tau_1, \tau_2, \ldots, \tau_k$, and the censoring values $r_1, \ldots, r_k$. Subsequent work along these lines would also involve a design of the "testing pattern" for step-stress tests.

Note that testing under a very severe environmental condition, say S_{k+1}, where $S_k \prec S_{k+1}$, could make the realizations of $X_{k+1\bullet}$ degenerate at 0. That is, what would be observed under S_{k+1} are instantaneous failures. The recording of instantaneous failures occurs because of a finite resolution of the recording device, and should not be interpreted as $X_{k+1\bullet} \equiv 0$. Thus, in what follows, S_k is chosen to be the most severe environmental condition for which the realizations of $X_{k\bullet}$ are not identically zero.

2 OVERVIEW

The statistical model, which includes a choice of the parametric family of distributions for F and also a choice of the time transformation function is described in Section 3. A discussion of some consequences of the above choices is also included in Section 3. Section 4 describes the main contribution of the paper, namely the development of a paradigm for the conduct of an optimal accelerated life test under the statistical model of Section 3. It turns out that whereas a prescription of the needed paradigm can, in principle, be formulated in an appealing manner - via Shannon's measure of information - progress towards its implementation calls for a consideration of simplifying scenarios and special cases. A simplifying scenario is outlined in Section 4, and a consideration of some special cases given in Section 5. These indicate the flavour of the approach and provide insights about the more general situations - insights which are otherwise counter intuitive. The hope here is that continued activities along the suggested lines will lead to developments which will contribute towards the generation of a comprehensive package of coherent techniques which will help address the practical issues raised by this important problem. These and other such matters constitute Section 6, which concludes the paper.

3 THE STATISTICAL MODEL

Suppose that X_{ij} is lognormal with parameters μ_i and σ_i^2, for $1 \leq i \leq k$, and $1 \leq j \leq r_i$. This assumption is not atypical in life testing; furthermore, it facilitates connection with the well developed Bayesian theory of linear normal models. For the time transformation function, the **Power Law**, popular in both biometry and reliability [Sethuraman and Singpurwalla (1982)] is assumed. Under this law, two versions of the time transformation function appear to be reasonable, one in terms of $E(X_{ij})$, and the other in terms of $M(X_{ij})$, the median of X_{ij}. Specifically, for

unknown constants $C > 0$ and $P \geq 0$, it is assumed that

$$E(X_{ij}) = \exp(\mu_i + \frac{\sigma_i^2}{2}) = \frac{C}{S_i^P}, \tag{1}$$

or that

$$M(X_{ij}) = \exp(\mu_i) = \frac{c}{S_i^P}, \tag{2}$$

If $Y_{ij} = \log X_{ij}$, then X_{ij} having a lognormal distribution with parameters μ_i and σ_i^2 - henceforth written as "$X_{ij} \sim \Lambda(\mu_i, \sigma_i^2)$" - implies that Y_{ij} is normal with mean μ_i and variance σ_i^2 - henceforth written as "$Y_{ij} \sim N(\mu_i, \sigma_i^2)$". As a **model of observations**, it is common to suppose that

$$Y_{ij} = \mu_i + \epsilon_{ij}, \text{ where } \epsilon_{ij} \sim N(0, \sigma_i^2); \tag{3}$$

then, letting $a = \log C, b = -P, V_i = \log S_i$ and taking logarithms in (1) and (2), the above relationships may be re-written, under (1), as

$$Y_{ij} = a + bV_i - \frac{\sigma_i^2}{2} + \epsilon_{ij}, \tag{4}$$

and

$$Y_{ij} = a + bV_i + \epsilon_{ij}\ , \tag{5}$$

under (2). It is instructive to note that under (4) the model of observations for the actual life lengths takes the form $X_{ij} = E(X_{ij})\tilde{\epsilon}_{ij}$, and under (5) the form $X_{ij} = M(X_{ij})\tilde{\epsilon}_{ij}$, where $\tilde{\epsilon}_{ij} = \exp(\epsilon_{ij})$; that is $\tilde{\epsilon}_{ij} \sim \Lambda(0, \sigma_i^2)$. Such relationships imply that an observed life length is its mean (median) multiplied by an innovation from lognormal distribution with median 1 and mean $\exp(\sigma_i^2/2)$. Furthermore, when σ_i^2 is assumed and $r_i = n_i$ for all i known, the relationships (4) and (5) imply

that the response vector $\mathbf{Y}$, where $\mathbf{Y} = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, \ldots, Y_{k1}, \ldots, Y_{kn_k})^T$, has the structure of a linear model of the form

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{6}$$

where, $\boldsymbol{\epsilon} \sim N(0, \Sigma), \boldsymbol{\beta}^T = (a, b)$. The design matrix, $\mathbf{A}$, is given by

$$\mathbf{A}^T = \begin{bmatrix} 1 \ldots 1 & 1 \ldots 1 & \ldots & 1 \ldots 1 \\ \underbrace{V_1 \ldots V_1}_{n_1} & \underbrace{V_2 \ldots V_2}_{n_2} & \ldots & \underbrace{V_k \ldots V_k}_{n_k} \end{bmatrix},$$

and Σ is the specified covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 I_{n_1} & & & \mathbf{0} \\ & \sigma_2^2 I_{n_2} & & \\ \mathbf{0} & & \ddots & \\ & & & \sigma_k^2 I_{n_k} \end{bmatrix}$$

where I_{n_j} is the $n_j \times n_j$ identity matrix.

Since Σ is assumed known, there are several strategies that could be used to facilitate a specification of the σ_i^2, all consistent with the physical aspects of the life testing scenario. In particular, it is reasonable to expect that σ_i^2 decreases with S_i, so that $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_k^2$. With regard to the above, several possibilities come to mind, one of which is that $\sigma_i^2 = K/S_i$, where K is specified, or that $\sigma_i = \sigma^2/i$, where σ^2 is specified.

In view of the above considerations and (6), it follows that

$$(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{\beta}, \Sigma) \sim N(\boldsymbol{A\beta}, \Sigma). \tag{7}$$

A final ingredient that remains to be specified is a prior assumption for the vector $\boldsymbol{\beta}$. For this, it is reasonable to assume that

$$(\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \Sigma_0) \sim N(\boldsymbol{\beta}_0, \Sigma_0), \tag{8}$$

where

$$\boldsymbol{\beta}_0^T = (a_0, b_0), \text{ and } \Sigma_0 = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$$

are specified. In cases where P is close to zero and $P \geq 0$, the assumption that $b = -P$ is normally distributed may be disturbing. A strategy for overcoming this in this scenario is to make σ_b^2 very small.

A generalization of (8) would involve a distributional assumption on $\boldsymbol{\beta}_0$ leading to the hierarchical model of Lindley and Smith (1972). Such a generalization would be prompted by concerns about the validity of the power law as a time transformation function, or about the appropriateness of the time transformation function over the entire range of stresses; see for example, Blackwell and Singpurwalla (1988). For the purposes of this paper we proceed with assumption (8).

It is well known [Lindley and Smith (1972)], that under (7) and (8), the posterior distribution of $\boldsymbol{\beta}$, given the data $\boldsymbol{y}$ obtained under $\boldsymbol{A}$, is of the form

$$(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{A}, \Sigma, \boldsymbol{\beta}_0, \Sigma_0) \sim N(\boldsymbol{Dd}, \boldsymbol{D}), \tag{9}$$

where

$$\boldsymbol{D} = (\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A} + \Sigma_0^{-1})^{-1} \text{ and } \boldsymbol{d} = (\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{y} + \Sigma_0^{-1}\beta_0).$$

Here $\boldsymbol{y}$ is a realization of $\boldsymbol{Y}$ and is obtained by replacing each Y_{ij} in $\boldsymbol{Y}$ by its realization y_{ij}. The use of (9) in designing an optimal accelerated test is explicated in the next section.

4 DESIGN OF AN OPTIMAL TEST

It was stated that the aim of accelerated testing is to facilitate inference about $Y_u = \log X_u$. Variables which control the quality of inference are the design matrix $\boldsymbol{A}$, the censoring values $r_1 \ldots, r_k$, and the test truncation times $\tau_1 \ldots, \tau_k$. Factors which dictate choices of the above variables are the costs of testing - where costs would include the time spent waiting for the results of the test. Optimal choices of the design variables would involve a trade-off between the costs of testing and the quality of the inference. Thus for example, testing at V_u itself (or as close to it as is possible) would not only be time consuming but would also result in data with a large variability; testing as far away from V_u as is possible would be expeditious and produce data with small variability. However, testing at very large values away from V_u could result in recording and instantaneous failure of all items on test, providing little, if any, information about Y_u. Advantages of testing in the vicinity of V_u are a savings in costs due to the non-failure of some of the items on test, and the knowledge that the observed data are likely to be more representative of the actual life lengths under V_u, than data that would be obtained when testing away from V_u. This latter issue would be germane if there were concerns associated with the validity of the time transformation function over the range of stresses considered. The set-up of Section 3 assumes that the power law is operative over the entire range of values of the stresses, and this diminishes the sense of the above argument for testing in the vicinity of V_u.

4.1 Consideration of a Simplifying Scenario

In many scientific experiments, particularly those involving biomedical scenarios, the cost of testing is either secondary, as compared with the quality of inference, or is the same at all test levels; thus in what follows the costs of testing will not be considered. Furthermore, for the purpose of illustration, it will be assumed that $r_i = n_i$, and that the τ_i's are infinite for all i. These assumptions imply no form of censoring and truncation, leaving an optimal choice of the design matrix $\boldsymbol{A}$ as the only decision variable. It will be shown in Section 4.3 that an optimal choice of $\boldsymbol{A}$ is facilitated by a consideration of the expected gain in Shannon information, for which the prior and posterior predictive distributions of Y_u are the essential ingredients.

4.2 The Predictive Distributions

Under the model of Section 3, it is clear that

$$(Y_u \mid V_u, \sigma_u^2, \boldsymbol{\beta}) \sim N([1\ V_u]\boldsymbol{\beta}, \sigma_u^2),$$

and we require the following predictive distributions; $Y_u|V_u, \sigma_u^2, \boldsymbol{\beta}_0, \Sigma_0$ and $Y_u|V_u, \sigma_u^2, \Sigma, \mathbf{y}, \boldsymbol{\beta}_0, \Sigma_0$. It can be shown that

$$(Y_u \mid V_u, \sigma_u^2, \boldsymbol{\beta}_0, \Sigma_0) \sim N\left([1\ V_u]\boldsymbol{\beta}_0, [1\ V_u]\Sigma_0 \begin{bmatrix} 1 \\ V_u \end{bmatrix} + \sigma_u^2\right), \tag{12}$$

and that

$$(Y_u \mid V_u, \sigma_u^2, \boldsymbol{\beta}_0, \Sigma_0, \Sigma, \boldsymbol{y}) \sim N\left([1\ V_u]\boldsymbol{D}\boldsymbol{d}, [1\ V_u]\boldsymbol{D} \begin{bmatrix} 1 \\ V_u \end{bmatrix} + \sigma_u^2\right), \tag{13}$$

4.3 The Design Criterion

The criterion used here to obtain the optimal design is to choose that design matrix $\boldsymbol{A}$ which maximizes the expected gain in Shannon information. Lindley (1956) proposed the use of Shannon information gain between prior and posterior in the context of Bayesian inference. Stone (1959a) explored this further in the content of experimental design. Stone (1959b) discussed its use in the Bayes hierarchical modeling framework as a unifying mechanism for previous *ad hoc* classical criteria. Smith and Verdinelli (1980), Kadane and Verdinelli (1990) discuss the application of Shannon information within the Bayes hierarchical modelling framework. Whittle (1973) discusses computational techniques for the computation of the optimal design. Chaloner and Larntz (1986) apply these techniques to nonlinear models.

Specifically, in our case, one needs to maximize, with respect to $\boldsymbol{A}$, the quantity

$$E_{\boldsymbol{y}} \int p(y_u \mid \boldsymbol{y}) \log \left(\frac{p(y_u \mid \boldsymbol{y})}{p(y_u)} \right) dy_u, \tag{14}$$

where $E_{\boldsymbol{y}}$ denotes expectation with respect to the marginal distribution of $\boldsymbol{Y}$. This marginal distribution depends on $\boldsymbol{A}, \boldsymbol{\beta}_0, \Sigma_0$ and Σ, and is given [Lindley and Smith (1972)] as

$$(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{\beta}_0, \Sigma_0, \Sigma) \sim N(\boldsymbol{A}\boldsymbol{\beta}_0, \boldsymbol{A}\Sigma_0\boldsymbol{A}^T + \Sigma). \tag{15}$$

The integral in (14) is the Kullback-Leibler divergence between the two distributions (12) and (13). Note that if $Z_l \sim N(m_l, s_l^2), l = 1, 2$, then the Kullback-Leibler divergence between the distributions of Z_1, and Z_2 is given by

$$KL(Z_1, Z_2) = -\frac{1}{2} - \frac{1}{2} \log \left(\frac{s_2^2}{s_1^2} \right) + \frac{1}{2} \left[\frac{s_2^2}{s_1^2} + \frac{(m_1 - m_2)^2}{s_1^2} \right] .$$

The m_1 and s_1 associated with (12) are

$$m_1 = [1\ V_u]\beta_0\ ,\ s_1^2 = [1\ V_u]\Sigma_0 \begin{bmatrix} 1 \\ V_u \end{bmatrix} + \sigma_u^2\ ,$$

and these do not depend on $\boldsymbol{y}$ nor on $\boldsymbol{A}$. However, the m_2 and s_2 associated with (13) depend on both $\boldsymbol{y}$ and and $\boldsymbol{A}$, respectively, and can be written as

$$m_2 = [1\ V_u][\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A} + \Sigma_0^{-1}]^{-1}[\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{y} + \Sigma_0^{-1}\beta_0],$$

and

$$s_2^2 = [1\ V_u][\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A} + \Sigma_0^{-1}] \begin{bmatrix} 1 \\ V_u \end{bmatrix} + \sigma_u^2.$$

Substituting the above in (14) results in the need to find that $\boldsymbol{A}$ for which the following expression is maximized

$$-\log t(\mathbf{A}, V_u) + t(\mathbf{A}, V_u) + E_{\boldsymbol{y}}(m_1 - m_2)^2. \tag{16}$$

where,

$$t(\mathbf{A}, V_u) = [1\ V_u][\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A} + \Sigma_0^{-1}]^{-1} \begin{bmatrix} 1 \\ V_u \end{bmatrix} + \sigma_u^2\ .$$

Clearly, an anlytical implementation of the above maximization poses a formidable task. However, a simplification of some of the terms in the above results in expressions which facilitate numerical computations. Specifically,

$$E_{\boldsymbol{y}}(m_1 - m_2)^2 = \text{Trace}\left\{ \text{Var}(\boldsymbol{Y})\Sigma^{-1}\boldsymbol{A}\boldsymbol{D} \begin{bmatrix} 1 & V_u \\ V_u & V_u^2 \end{bmatrix} \boldsymbol{D}\boldsymbol{A}^T\Sigma^{-1} \right\},$$

where $\mathrm{Var}(\boldsymbol{Y}) = \boldsymbol{A}\Sigma_0\boldsymbol{A}^T + \Sigma$ from (15). Furthermore,

$$[\boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A} + \Sigma_0^{-1}]^{-1} = \begin{bmatrix} \sum_{i=1}^k \frac{n_i}{\sigma_i^2} & \sum_{i=1}^k \frac{V_i M_i}{\sigma_i^2} \\ \sum_{i=1}^k \frac{V_i M_i}{\sigma_i^2} & \sum_{i=1}^k \frac{V_i^2 M_i}{\sigma_i^2} \end{bmatrix} + \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}^{-1},$$

and the ingredients necessary to obtain $\boldsymbol{A}\Sigma_0\boldsymbol{A}^T + \Sigma$ are specified in Section 3.

It is of interest to note that were inference about $\boldsymbol{\beta}$, rather than Y_u, the focus of attention, then $\boldsymbol{A}$ would be chosen so as to maximize, see Smith and Verdinelli (1980).

$$|\ \mathrm{Var}(\boldsymbol{\beta} \mid \boldsymbol{Y})\ |^{-1} = |\ \boldsymbol{A}^T\Sigma^{-1}\boldsymbol{A} + \Sigma_0\ |\ , \tag{17}$$

rather than the expression (16) given before. Furthermore if the cost of the testing is an issue of concern, then one needs to specify the value of information on a scale comparable with monetary units.

5 CONSIDERATION OF SPECIAL CASES

The simplest case to consider is the one which involves knowing C, say $C = 1$, so that $a_0 = 0$. If for all $i, i = 1, \ldots, k, \sigma_i^2 = n_i = 1$, and $\sigma_b^2 = k = 1$, that is only a single unit can be subjected to an accelerated test, then the optimal design problem finds the stress at which the item should be tested - in the vicinity of V_u, or at V_k, the largest stress. For this scenario $\Sigma = \Sigma_0 = 1$. Algebraic manipulation shows that the single item should be tested at V_k. This conclusion may be contrary to intuition, because many, influenced by the notion of a warm feeling, would test in the vicinity of V_u. Furthermore, even if more than one item were available for testing, one would test all the items at V_k, and the same would also be true if σ_i^2 were to be decreasing in i.

In connection with the above, an important question that needs to be addressed is that pertain-

ing to the number of items that should be tested at V_k; that is, how large should n_k be? For this, one investigates the behavior of (16) - the expected gain in Shannon information - as a function of n_k, and chooses that value of n_k after which (16) shows little or no improvement. For $\Sigma = \Sigma_0 = 1$, the behavior of (16) as a function of n_k, for $V_k = 2$, 4, 6 and 8 respectively, and $V_u \neq 0$, is shown in Figure 5.1. It is interesting to note that the optimum value of n_k decreases in V_k; this is to be expected since the variance of the observed life-lengths decreases as the stress increases.

When P is known, say $P = 1$, but C unknown, a situation quite common when dealing with the fatigue life of ball bearings [Ioannides and Harris (1985)] suggests that a transformation of the data would result in a model of the form $\tilde{Y}_{ij} = a + \epsilon_{ij}$, where $\tilde{Y}_{ij} = Y_{ij} - bV_i + \frac{\sigma_i^2}{2}$, under (2) and $\tilde{Y}_{ij} = Y_{ij} - bV_i$, under (5). Clearly, if σ_i^2 decreases in i - as is usually assumed, then one would test all the items under stress V_k where the variance of ϵ_{kj} is the smallest. Were σ_i^2 assumed to be a constant for all values of i, then it would not matter at which stress the items are tested. The above conclusion can also be arrived at by proceeding formally in terms of maximizing (14). Finally, the determination of an optimal n_k would be based on a plot of the behavior of the utility versus n_k, similar to that of Figure 5.1.

When both C and P are unknown - as is often the case - and when only one item is available for testing, then under the assumption that $\Sigma = 1$ and $\Sigma_0 = I_2$, it can be shown that the optimum stress V^* at which testing is to be done is given by that value of V, V^* say, which attains

$$\max_V \left(-\log t(V) + t(V) + \frac{(1 + V_u V)^2}{V^2 + 2} \right) \tag{18}$$

where $t(V) = \frac{1}{V+2}\left((V - V_u)^2 + 1\right)) + \sigma_u^2$. Interesting, the answer depends on V_u. If $V_u = 0$ (that is, $S_u = 1$), then V^* turns out to be V_u.

In figure 5.2, we show a plot of the expected gain in Shannon information as a function of V_i,

when $V_u = 0$. Observe that the expected gain in Shannon information is a monotonically decreasing function of V_i implying that when $\Sigma = 1$, it is optimum to test at $V_u = 0$.

Suppose now, that in the scenario given above, testing at V_u is not possible and that one could test anywhere in the range $V_1 = 1$ through $V_k = 6$. Suppose the experimenter has two observations with one at $V = 2$. He wishes to choose the position of the second one optimally. In Figure 5.3, we show a plot of the expected gain in Shannon information as a function of $V)i$, the stress for the second item. It is interesting to note that testing at $V_k = 6$ would be preferable to testing at any $V_i > 1.25$, but that testing in the range $(1, 1.25)$ is preferable to $V_k = 6$.

If $\sum = \sigma_i^2$, where σ_i^2 known but decreasing in i, then, from (18), the choice V^* would be a trade-off between the "loss in utility " due to a large variance for testing in the vicinity of V_u, versus the "loss in utility" due to our lack of knowledge about C and P which gets emphasized when we test away from V_u.

Our development leads to some interesting conclusions when both C and P are unknown and when we are allowed to test 2 items. Under the assumption that $\Sigma = \Sigma_0 = I_2, \sigma_u^2 = 1$, $V_u = 0$ and testing is possible at $V_u = 0$, then the optimum decision would be to test <u>both</u> the items at V_u; however, if testing at V_u were not possible and if $V_1 = 2$ is the weakest stress at which testing could be done is $V = 2$, then the optimum decision would be to test one item at $V_1 = 2$ and the other at $V_k = 6$. Furthermore, if $V_1 = 1$ and $V_k = 5$, then the optimum decision would be to test both the items at $V_1 = 1$, or to test one item at $V_1 = 1$ and the other at $V_k = 5$, both scenarios giving the same Shannon information gain.

The implication of the above conclusions is that to maximize the expected gain in Shannon information, as defined by (14), one should test, if possible, at or in the vicinity of V_u to gain direct knowledge about Y_u. However, if testing in the vicinity of V_u is not possible, then one should gain knowledge about Y_u via a knowledge about C and P, and the latter is optimally obtained by

spreading out the test points as far apart as is possible. It so happens that when $V_1 = 1$ and $V_k = 5$, it does not matter how knowledge about Y_u is obtained - by direct observation in the vicinity of V_u or by knowledge about C and P.

6 CONCLUSIONS

It is evident from the material of Section 5 that the development given here enables one to address many of the practical issues posed by the problem of designing accelerated tests. The special cases of Section 5 have assumed that σ_i^2 is a constant. We have focused on the unknowns in the time transformation function. Clearly, an important parameter to learn about is σ_u^2 itself. Furthermore, we have not considered the case of time truncation and censoring, nor have we addressed the issue of choosing optimal sample sizes. However, all of the above considerations can in principle be undertaken - the only limitation being a question of details. All the same, our approach does bring out matters that are not otherwise intuitive and the special cases give us many clues about what to expect and what direction to proceed when considering the more general scenarios.

Typically, in linear model with k parameters, designs concentrate on k points. This is not generally true in nonlinear situations (Chaloner and Larntz, 1986). In our case we have $k = 2$. Stone (1959) considers the number of points required to maximise Shannon information when estimation is the aim. The extension to the predictive case requires future research.

APPENDIX

Suppose that the experimenter is faced with a decision problem for the random variable Y_u, with realisation, y_u. Let $U(y_u, d)$ denote the utility function, where $d \in D$, the decision space.

Define the information gained, $E(I(Y))$ through observing the random variable Y by

$$E(I(Y)) = E_Y\left(\sup_{d\in D} E_{Y_u|Y}\left(U(Y_u,d)\right)\right) - \sup_{d\in D} E_{Y_u}\left(U(Y_u,d)\right) . \tag{19}$$

If D is the space of probability distributions for prediction at the point V_u and the utility function, for reporting probability densities, is the logarthimic function, that is, $U(y_u, d(\cdot)) = \log d(y_u)$, then,

$$\sup_{d\in D} E_{Y_u|Y}\left(U(Y_u,d)\right) = \int p(y_u|Y)\log p(y_u|Y)dy_u$$

and

$$\sup_{d\in D} E_{Y_u}\left(U(Y_u,d)\right) = \int p(y_u)\log p(y_u)dy_u .$$

Hence the information gain is

$$E(I(Y)) = E_{(Y,Y_u)}\left[\log\left(\frac{p(Y_u|Y)}{p(Y_u)}\right)\right]$$

which is the Shannon information between the posterior predictive density at Y_u to the prior predictive density at Y_u. If estimation under a quadratic loss function is the aim then the criteria of maximising information gained becomes that of minimising posterior variance at the point V_u, that is criteria (17).

References

[1] BLACKWELL, L.M. and SINGPURWALLA, N.D. (1988). Inference from accelerated life tests using filtering in coloured noise. *J. R. Statist. Soc. B.*, **50**, No. 2, 281-292.

[2] CHALONER, K. and LARNTZ, K. (1986). Optimal Bayesian design applied to logistic regression

experiments. *Technical Report 483*, University of Minnesota.

[3] DeGroot, M.H. and Goel, P. (1979). Bayesian estimation and optimal designs in partially accelerated life testing. *Naval Res. Logistics quarterly*, **26**, 223-235.

[4] Ioannides, E. and Harris, T.A. (1985). A new fatigue life model for roller bearings. *Journal of Tribology*, **107**, 367-378.

[5] Kadane, J.B. and Verdinelli, I. (1990). Bayesian designs for maximizing information and outcome. *Technical Report* 461, Carnegie Mellon University.

[6] Lindley, D. V. (1956). On the measure of information provided by an experiment. *Ann. Statist.*, **27**, 986-1005.

[7] Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). *J. R. Statist. Soc. B.*, **34**, 1-41.

[8] Mann, Schafer and Singpurwalla (1974). Methods for statistical analysis of reliability and life data. John Wiley and Sons, Inc.

[9] Meinhold, R. and Singpurwalla, N.D. (1987). A Kalman filter smoothing approach for extrapoletions in certain dose-response, damage assessment and accelerated life testing studies. *Amer. Statist.*, **41**, 101-106.

[10] Sethuraman, J.S. and Singpurwalla, N.D. (1982). Testing of hypotheses for distributions in accelerated testing. *J. Amer. Stat. Assoc.*, **77**, 204-208.

[11] Shaked, M. and Singpurwalla, N.D. (1983). Inference from step-stress accelerated tests. *Journal of Statistical Planning and Inference*, **7**, 295-306.

[12] SMITH, A. F. M. and VERDINELLI, I. (1980). A note on Bayes designs for inference using a heirarchical linear model. *Biometrika*, **67**, 613-619.

[13] STONE, M. (1959a). Application of a measure of information to the design and comparison of regression experiments. *Ann. Statist.*, **21**, 55-70.

[14] STONE, M. (1959b). Discussion of Kiefer, *J.R. Statist. Soc. B*, **21**, 313-315.

[15] WHITTLE, P. (1973). Some general points in the theory of optimum experimental design. *J.R. Statist. Soc. B*, **35**, 123-130.

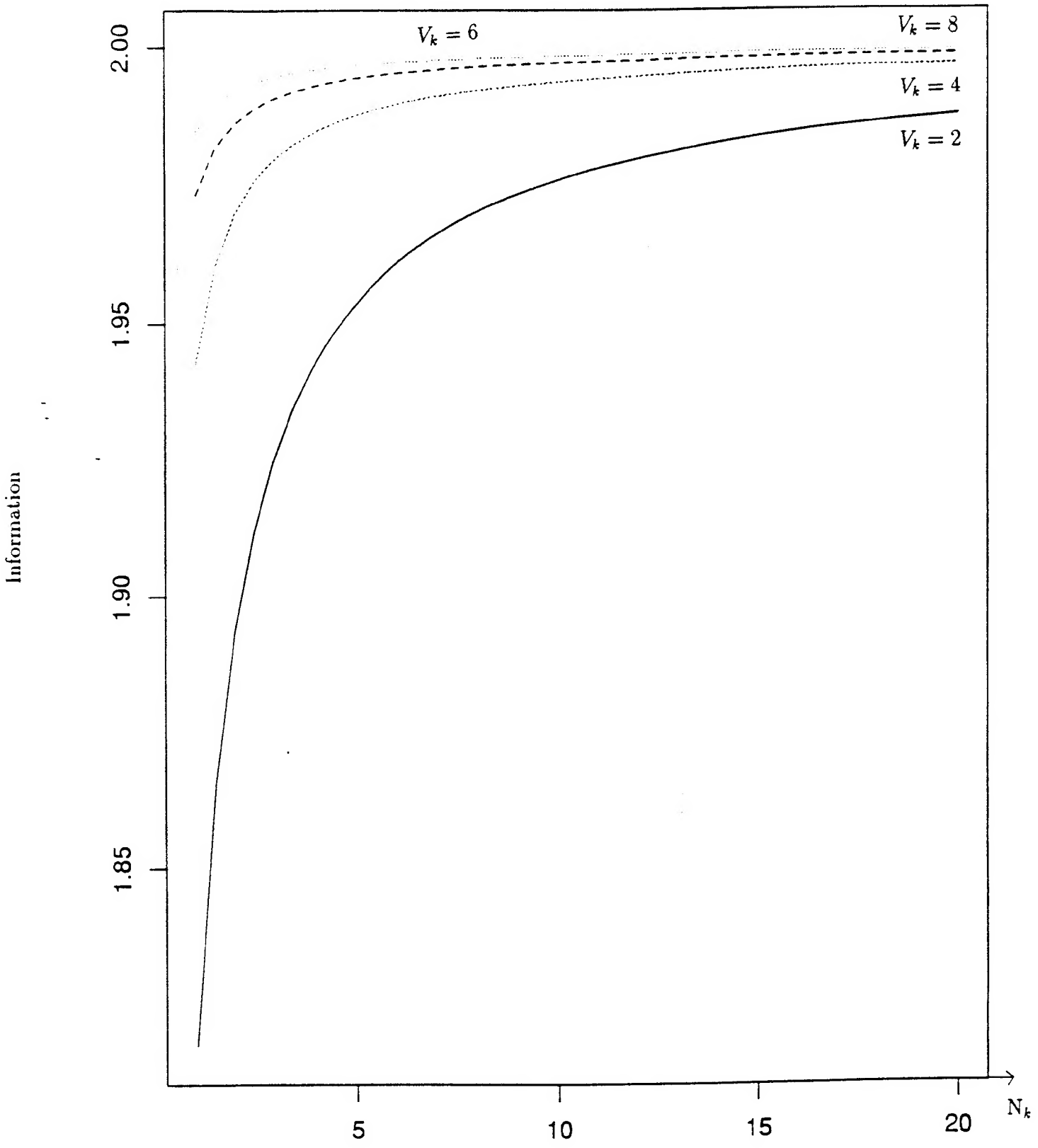


Figure 5.1. Expected Gain in Shannon Information as a Function of N_k, for $V_k = 2, 4, 6$ and 8, $V_u \neq 0$ and C is known.

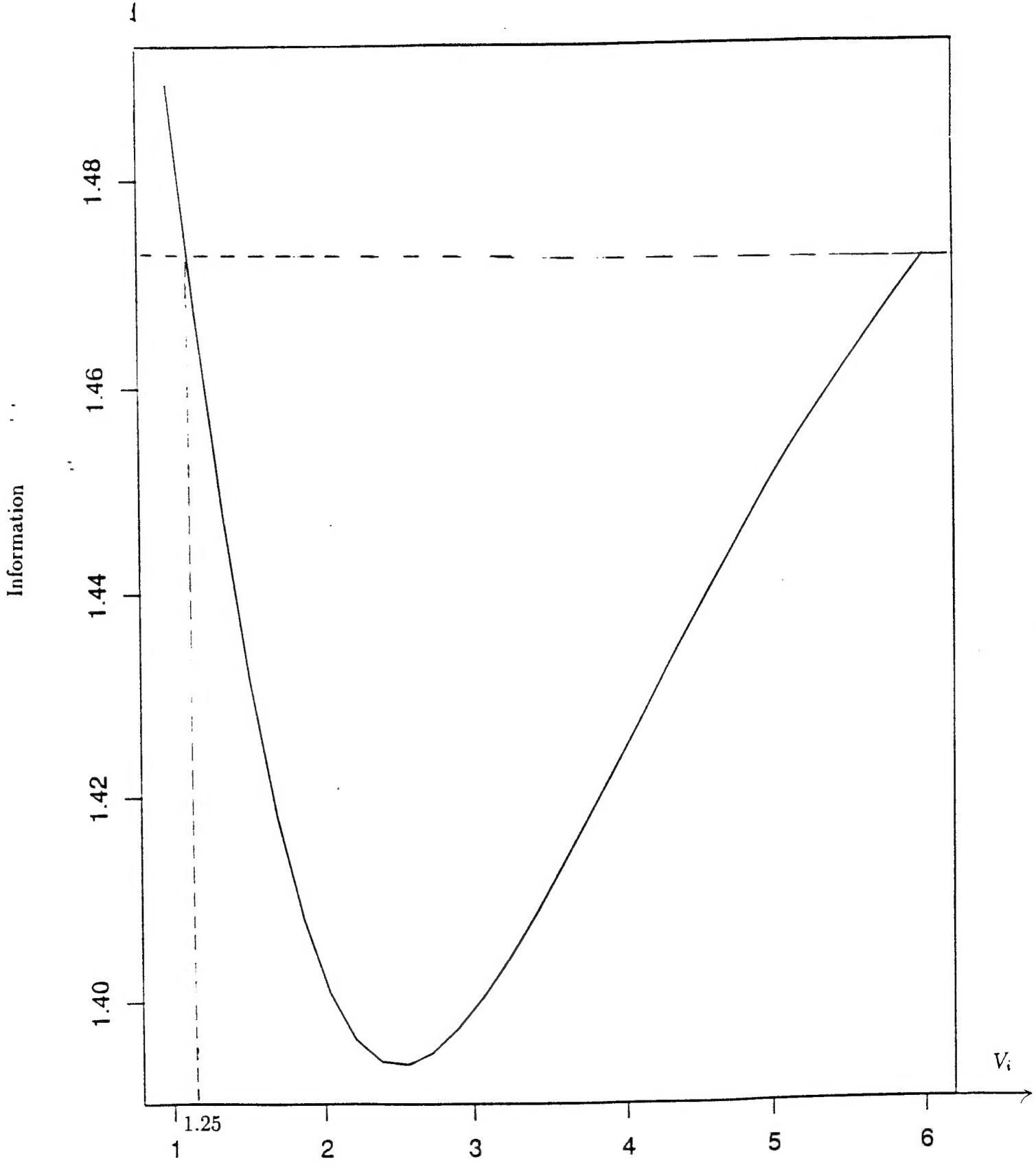


Figure 5.3. Expected Gain in Shannon Information as Function of V_i, Under the Conditions of Figure 5.2 but when $1 \leq V_i \leq 6$

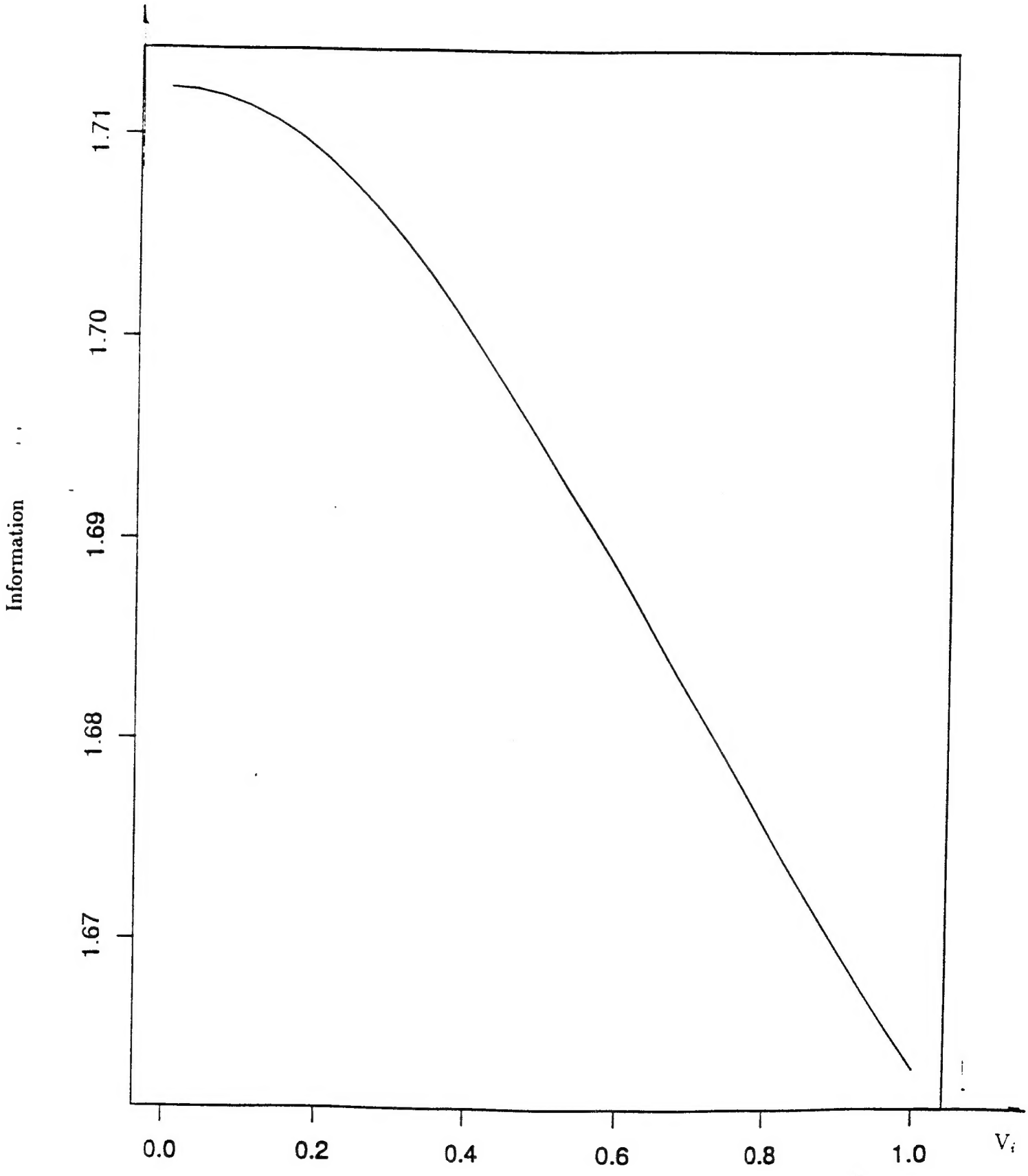


Figure 5.2. Expected Gain in Shannon Information as a Function of V_i the Test Stress when $V_c = 0$, $\Sigma = n = 1$ and C and R are values